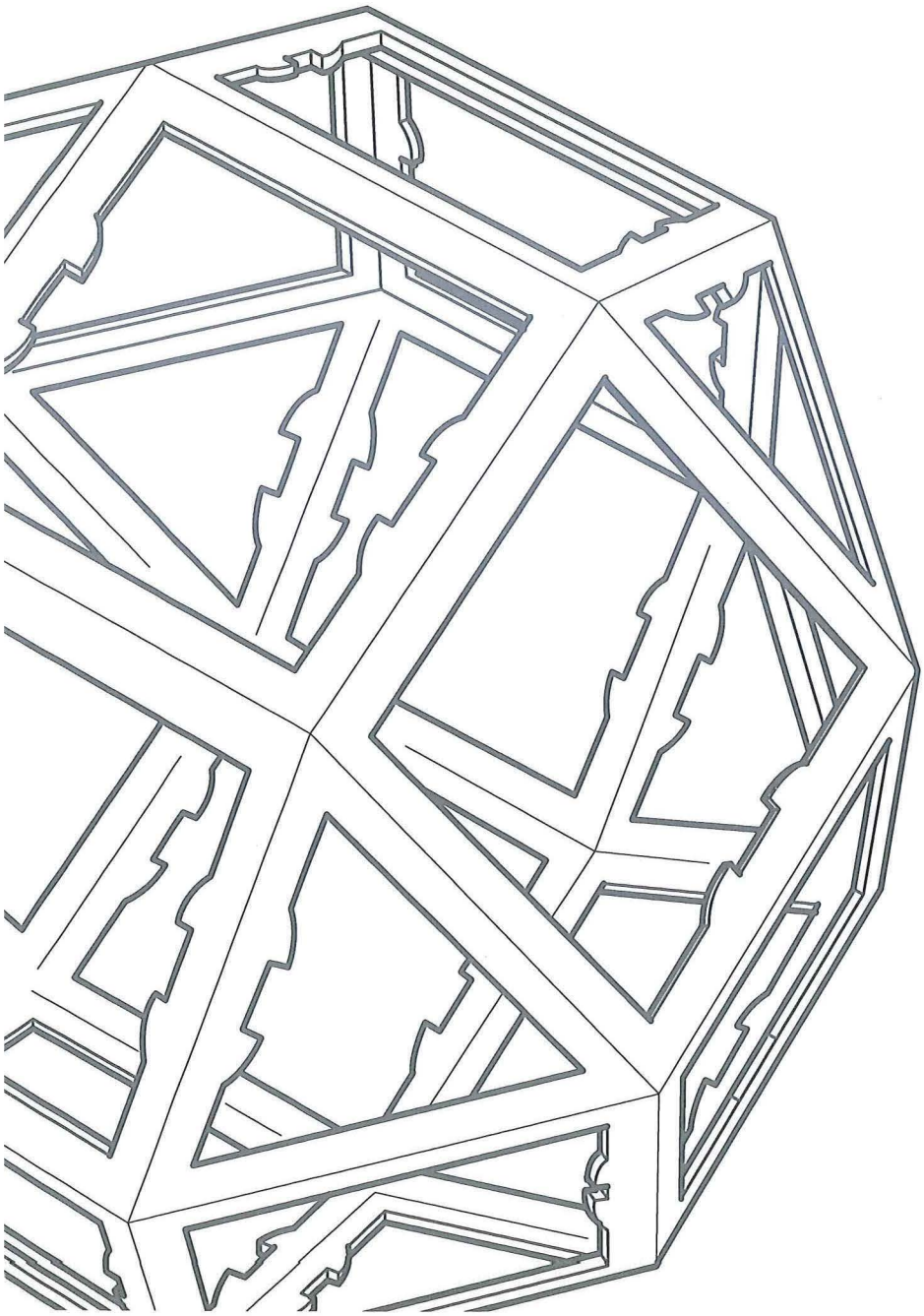
Department of Mathematics and Statistics
College of Engineering

UC
UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Summer Research Project

Generalised Linear Mixed Models and its Application in R

By James Dawber



09

Generalised Linear Mixed Models and its Applications in R

James Dawber

February 2009

Abstract

Through the benefits of mixed modeling over the usual fixed effects modeling, certain data sets can be better understood. Repeated measures data is especially well suited to mixed models. The repeated measures can be used as random effects and thus improve the modeling process through the mixed model. Statistical software such as R is now becoming increasingly more useful in using these mixed models. Through using R, repeated measures data is readily modeled using generalised linear mixed models. There are two packages in R which can perform generalised linear mixed models. The application of both these packages in modeling generalised linear mixed models is explained simply and concisely, with straightforward guidelines given to assist in the modeling process. Comparisons are made between the modeling methods within each of the two packages, with both the benefits and limitations of each package highlighted. Methods for additional analysis on these models are also described. Repeated measures count data on New Zealand birds is used as an example to show exactly how to implement generalised linear mixed models in both packages in R.

Introduction

Data collected in certain experiments will often contain observations that are clustered, correlated or somehow dependent on each other. For example, this could be due to individuals being related (perhaps students at the same school) or individuals being observed regularly over time as part of a longitudinal study. Response vectors that have correlated observations can be modeled using a mixed model (Fox, 2002). Mixed models have additional random effects terms which are appropriate for representing these correlated observations. Statistical software such as R can be now used to apply mixed models to such data. There is in fact, two packages in R which can perform mixed modeling. The following example on bird counts will be used as a basis in which to simply model both linear mixed models and the generalised form (with non-normally distributed data) in both these packages in R. The aim is to give the reader simple and methodological instructions to implement mixed models in R.

Example: Bird Counts

Ongoing monitoring of bird numbers associated with a habitat restoration project was initiated in 2004. Every year since 2004, the number of each different species was counted in the treatment area, and in two non-treatment areas. Twelve transects were randomly placed in the treatment area and six each of the two non-treatment areas. To get an accurate estimate of the number of birds within each year, each transect was counted five times per day over three separate days of a week. It is essential to find out the trend of bird numbers over the five years and whether there is a difference between the treatment and non-treatment areas. The best way to model such longitudinal data is to use mixed models as observations within each transect, day or week will be likely to be correlated. Through this analysis the effects of the habitat restoration project on the bird species will be identified and better understood.

Background

Linear Mixed Models

Typically, a simple linear regression model is expressed as

$$y = X\beta + \epsilon \tag{1}$$

which is composed of a response vector y, a matrix X of the explanatory variables, a vector of unknown coefficient parameters β and a vector of unknown unobservable random errors ϵ. In this case, the explanatory variables are said to be fixed effects. These fixed effects are usually associated with all individuals of interest. However, there occur times when it may seem logical to assume these variables are not fixed, but random effects. Random effects are usually associated with individual units being drawn out randomly from a population (McCulloch and Searle, 2001). The random effects model naturally occurring deviations between the specified individual units as a part of some known underlying distribution (usually the normal distribution). Random effects variables can correspond to different levels, factors, subjects or experimental units in the model that are often correlated (Jiang, 2007) . For our example, a selection of transects were randomly laid in each of the areas, so the transects will contribute random effects on the model. In general, a variable is a random effect when the interest is not on the individual levels themselves, but the population that the levels represent (Faraway, 2006). For example, we are not necessarily interested in the individual transects but the infinite population of transects which makes up the area. In other examples it is difficult to establish whether the variable is a fixed or a random effect. Distinguishing between the two can be confusing as there are varying definitions of the terms across statistical literature (Gelman and Hill, 2007). It is advised to judge each model on its own merits to best decide which variables are fixed effects, and which are random effects. A model with both fixed and random effects is known as a mixed model.

A linear mixed model is expressed the same as a simple linear regression model, but with the additional random effects, typically given by

$$y = X\beta + Z\alpha + \epsilon \tag{2}$$

with a known matrix Z which contains categories attributing to the individual units and a vector of random effects α. Usually α is assumed to be normally distributed as $N(0, \sigma_R^2)$ but this is not a necessity, any distribution can be used. It is necessary for α to be uncorrelated with ϵ (Lee et al., 2006). It is also important that β and α are not correlated as this will induce bias in the model (Lee et al., 2006). In this simple case, the random effects will contribute towards a random intercept, however it is also common to have just a random slope or even to have both a random intercept and a random slope as shown in Figure 1.

The addition of random effects to a model reduces the unknown variation and therefore improves the entire model. Mixed modeling is especially useful when it comes to longitudinal and spatial data which often have correlated individual observations inside certain levels.

Generalised Linear Mixed Models

Just as generalized linear models can be applied to the typical fixed effects model, generalized linear mixed models (GLMM's) can be applied to mixed models. This is necessary as not all response variables are normally distributed, often they can be discrete or noticeably non-normal. GLMM's allow for the response data to come from any other distribution from the exponential family including most commonly the Binomial, Poisson and Gamma distributions.

Modeling the Data

The data is comprised of counts of birds across different areas, in different transects at different times. Because it is count data, the response variables can be assumed to come from a Poisson distribution. Each bird species is a different response variable. It is given in the data as *SpeciesTotal*, for example bellbirds are given as *BellbirdTotal*. The fixed effects are *Year*, the year in which the observation was made and *Treatment*, a categorical variable determining whether the observation is from the treatment area or not. The random effects are *Transect*, indicating which transect the observation came from, *Count*, the five daily repeats and *Rep*, the three separate day repeats. Some of the data is shown below:

```
> birds<-read.table("birds.txt",header=T)
> birds[1:7,1:8]
  Year Time  Area Transect Count Rep BellbirdTotal TuiTotal ...
1 2004  858 1        1     1   0            1        0
```

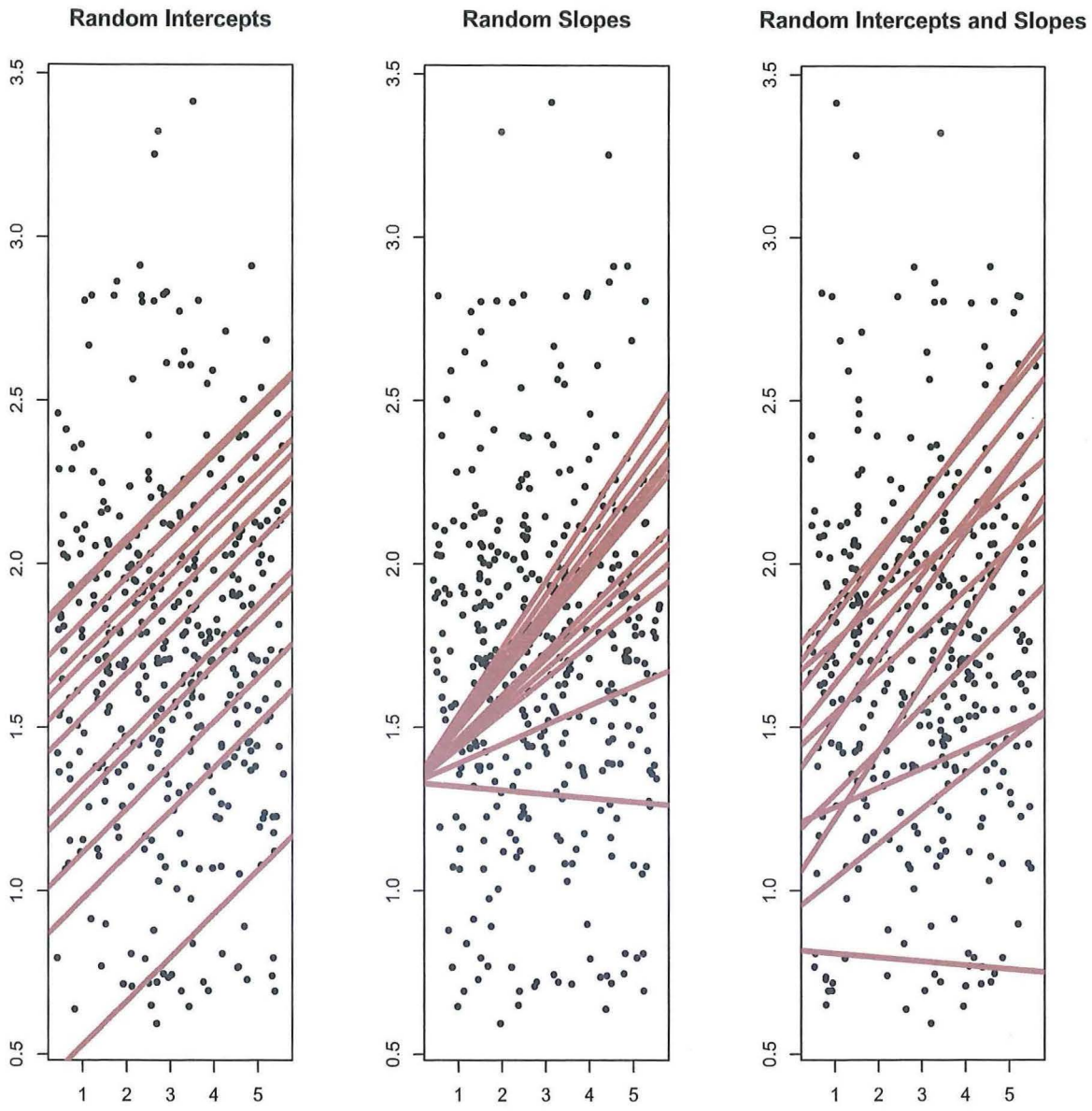


Figure 1: A plot showing differences in the applications of random effects models.

2	2004	912	1	1	2	0	2	0
3	2004	928	1	1	3	0	0	0
4	2004	946	1	1	4	0	1	0
5	2004	1003	1	1	5	0	1	0
6	2004	1204	1	2	1	0	3	1
7	2004	1232	1	2	2	0	2	0

We are modeling to see whether bird counts are changing over the years and between the treatments, thus we hold *Year* and *Treatment* as fixed effects as they are the specific variables of interest. This leaves the random effects *Transect*, *Count* and *Rep* which we are not specifically interested in. It is the area as a whole we are investigating, not the individual transects. We are interested in the variation within the day as a whole, not between the specific five *Counts*. Similarly, it is the variation within the week that is of interest, not between the three selected *Reps*. For these reasons, these three variables are considered as random effects. It can be assumed that the observations within the *Transect*, *Count* and *Rep* levels would perhaps show some kind of correlation. Having these variables as random effects acknowledges this correlation, and thus prevents possible violations to the assumptions of the model. So it is quite essential that they are used, as it improves the model greatly.

The random effects variables in this case demonstrate a hierarchy. Each transect is measured on each of the three days, and in each of the three days, five measurements are taken. This hierarchical formation arises when a variable is a subset of another variable. The *Transect* variable is of foremost importance, it is this variable which is supplying the main amount of information about the observations in the *Treatment*. The *Counts* and *Reps* are of less importance as they are merely repeats within each year that are not expected to create as much variation. It is clear that *Count* is a subset of *Rep*, that is, for every *Rep*, there are multiple *Counts*. *Count* is said to be nested within *Rep*. So the hierarchy in this model is *Count* nested in *Rep* which is nested in *Transect*. We will apply to the model random effects contributing to a random intercept and assume these random effects all follow a normal distribution.

Mixed Modeling and Computer Software

Deriving the estimates of β and α in mixed models involves complicated and timely calculations. For this reason, major developments in the field could not be easily made in the past without the advancement of modern computers. In the past two decades mixed modeling techniques have grown greatly due to the constant development of computers and the necessary computer software. Amongst the forerunners in computer software in the mixed model field is R.

R contains two packages which are made to run both linear mixed models and GLMM's. These two packages

are nlme and lme4. The first package to arise in R was the nlme package in 1999 which was very early on in R's expansion. The lme4 package arose later in 2003 by the co-author of nlme and is still in development. Although both packages perform the same tasks, there are substantial differences in how they run. The two packages both have their advantages and disadvantages over each other. When it comes to fitting mixed models in general, the lme4 package is recommended. Although the lme4 package is currently still in development, it is substantiating a very efficient and up-to-date system with fitting linear mixed models and GLMM's. The lme4 package should be the default package to use; however, it does sometimes have its disadvantages to the nlme package. The main advantage of this older package is that it has a more established grounding in using linear mixed models. It therefore has more tools available for interpreting and assessing the models which the lme4 package is yet to develop. Among these tools are `intervals()` which gives the confidence intervals of both the fixed effects and random effects coefficients for the model. Also, the authors of the package, Pinheiro and Bates, released a book called Mixed-Effects Models in S and S-PLUS which extensively and thoroughly covers the uses and theory behind the tools in the package. So when it comes to doing a linear mixed model and thorough analysis is required, it may be better to use the nlme package. However, when it comes to doing GLMM's, the nlme package is substandard to the lme4 package and has few benefits in this respect. For our example on birds, the lme4 package will be of primary focus, followed by the corresponding methods in the nlme package.

Example: Linear Mixed Modeling

Let us first assume that the response vectors in the example are not from a Poisson distribution, but from a normal distribution. And secondly, assume we just want to use *Transect* as the only random effect (i.e. discard *Rep* and *Count*). It is now unnecessary to apply a GLMM, but instead just a simple linear mixed model. To fit a linear mixed model in R, we first must ensure that all our random effects variables are factors, as well as any other discrete fixed effects variables. So we must turn *Treatment*, *Transect*, *Rep* and *Count* all into factors, done like so:

```
> attach(birds)
> Treatment<-Area==1
> Treatment<-factor(Treatment)
> fRep<-factor(Rep)
> fCount<-factor(Count)
> fTransect<-factor(Transect)
```

Note that treatment was not already a variable so it needed to be created. With all the variables created, the data can now be modeled. Let us look at how to model it in R by using the bellbird response as an example.

```
> library(lme4)
> lmm<-lmer(BellbirdTotal~I(Year-2003)*Treatment +(1|fTransect))
```

Firstly, the lme4 package is loaded. Then the function `lmer()` is used from the package to perform a linear mixed model. `lmer()` works similarly to the usual `lm()` function except the additional random effects term `(1|fTransect)` is added. In normal linear mixed models there are two established methods of modeling, there is maximum likelihood (ML) and restricted maximum likelihood (REML) (Jiang, 2007), both of which can be used in R when selected in the method argument. The relatively simple ML method generally gives biased variance component estimates however; henceforth REML is implicated as the default method on R. The REML method first transforms the data which in effect removes the fixed effects. From this, unbiased estimates can then be made about the unknown variance components (Jiang, 2007).

Of course, in our example *Transect* is not our only random effect, we have *Rep* nested in *Transect* and *Count* nested in *Rep*. This can simply be modeled with `lmer()` by altering the random effects term to `(1|fTransect/fRep/fCount)`. In general, if `a` and `b` are random effects and `a` is nested in `b` then it would be modeled as`(1|b/a)` using `lmer()`. Let us now carry out the full model and assess its results.

```
> lmm1<-lmer(BellbirdTotal~I(Year-2003)*Treatment + (1|fTransect/fRep/fCount))

> summary(lmm1)
Linear mixed-effects model fit by REML
Formula: BellbirdTotal ~ I(Year - 2003) * Treatment + (1 | fTransect/fRep/fCount)
  AIC  BIC logLik MLdeviance REMLdeviance
 5568 5606  -2777       5537         5554
Random effects:
 Groups                  Name        Variance  Std.Dev.
 fCount:(fRep:fTransect) (Intercept) 0.0518637 0.227736
 fRep:fTransect          (Intercept) 0.0044963 0.067055
 fTransect               (Intercept) 0.1880692 0.433670
 Residual                            1.2015126 1.096135
number of obs: 1800,
groups: fCount:(fRep:fTransect), 180; fRep:fTransect, 36; fTransect, 12

Fixed effects:
                          Estimate Std. Error t value
(Intercept)                1.59556    0.15306  10.424
I(Year - 2003)             0.02444    0.02584   0.946
```

```
TreatmentTRUE                -0.53611    0.12118  -4.424
I(Year - 2003):TreatmentTRUE  0.21833    0.03654   5.976

Correlation of Fixed Effects:
            (Intr) I(Yr-2003) TrTRUE
I(Yer-2003) -0.506
TretmntTRUE -0.396  0.640
I(Y-2003):T  0.358 -0.707     -0.905
```

There are four basic sections to the output. The first, displays the AIC, BIC and log likelihood which are essential in model comparisons. For the REML estimation, these model comparison tests will only work when the fixed effects structure are the same (Pinheiro and Bates, 2000).
The second section displays the variance (and standard deviation) components of the random effects and the error (residual) term. The third section displays the estimates of the coefficients for the fixed effects. The first thing you will realise is the lack of p-value; this is done for a reason. The author of the lme4 package, Douglas Bates, argues that certain assumptions are violated in the REML method and so it is therefore faulty to acquire p-values from the corresponding t-value. He suggests using the `mcmcsamp()` function to obtain p-values which uses Markov Chain Monte Carlo methods. You can then acquire confidence intervals for the fixed parameters like so;

```
> x<-mcmcsamp(lmm1,10000)   %Creates a sample of size 10000 from the posterior
                             distribution of the model's fitted parameters.
> library(coda)
> HPDinterval(x, prob=0.95) %Gives a 95% confidence interval for each parameter.
                                      lower         upper
(Intercept)                      1.28355527    1.93028576
I(Year - 2003)                  -0.02547958    0.07678482
TreatmentTRUE                   -0.76192976   -0.28614323
I(Year - 2003):TreatmentTRUE     0.14589810    0.28889740
log(sigma^2)                     0.11595364    0.25368676
log(fC:(.(In))                  -3.98457671   -2.25387045
log(fR:T.(In))                 -39.68289105   -3.24493476
log(fTrn.(In))                  -2.48703320   -0.61598806
deviance                      5537.78664264 5551.72886885
attr(,"Probability")
```

```
[1] 0.95
```

The fourth section shows the correlations between the fixed effects. As in any model, multicollinearity is a problem, linear mixed models are no different.

As mentioned earlier, p-values are not implemented in `lmer()` and this can be somewhat irritating if the modeler only requires a basic estimate of the p-value. If this is the case then using the nlme package would become beneficial as it does supply the p-value. The exact same model can be executed in nlme like so,

```
> library(nlme)
> lmm2<-lme(BellbirdTotal~I(Year-2003)*Treatment, random=~1|fTransect/fRep/fCount)
> summary(lmm2)
Linear mixed-effects model fit by REML
 Data: NULL
       AIC      BIC    logLik
  5569.817 5613.764 -2776.909
Random effects:
 Formula: ~1 | fTransect
        (Intercept)
StdDev:   0.4333816

 Formula: ~1 | fRep %in% fTransect
        (Intercept)
StdDev:  0.06749594

 Formula: ~1 | fCount %in% fRep %in% fTransect
        (Intercept) Residual
StdDev:   0.2278353 1.096114

Fixed effects: BellbirdTotal ~ I(Year - 2003) * Treatment
                                  Value  Std.Error   DF   t-value p-value
(Intercept)                   1.5955556 0.15299952 1617 10.428500  0.0000
I(Year - 2003)                0.0244444 0.02583565 1617  0.946152  0.3442
TreatmentTRUE                -0.5361111 0.12117995 1617 -4.424091  0.0000
I(Year - 2003):TreatmentTRUE  0.2183333 0.03653713 1617  5.975656  0.0000
 Correlation:
                              (Intr) I(Yr-2003) TrTRUE
```

```
I(Year - 2003)                -0.507
TreatmentTRUE                 -0.396  0.640
I(Year - 2003):TreatmentTRUE   0.358 -0.707     -0.905

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-2.38564243 -0.71382741 -0.04356208  0.62448070  4.04287664

Number of Observations: 1800
Number of Groups:
           fTransect        fRep %in% fTransect fCount %in% fRep %in% fTransect
                  12                          36                            180
```

It can now be easily seen which fixed effects are significant or not. It is important to remember that these p-values are believed to be breaching assumptions and caution should be taken when using them in practice.

So the model is showing that there is no evidence of a trend in bellbird numbers over the five years, however the treatment area is showing significantly less bellbirds. The interaction term is significant but is showing extremely high correlation (-0.905) with the treatment term indicating a somewhat ill-conditioned model (Fox, 2002). To emphasise, modeling this data assuming the response vector is normal is wrong, it is merely shown here as an example of how to use linear mixed models. It is appropriate to model the data as a GLMM.

Example: Generalised Linear Mixed Modeling

Modeling GLMM's in R is very similar to linear mixed models. Let us model the data in exactly the same way as above with `lmer()`, except now treating the response vector correctly as a Poisson random variable.

```
> glmm<-lmer(BellbirdTotal~I(Year-2003)*Treatment + (1|fTransect/fRep/fCount),
        family = poisson)
```

The only necessary thing to do differently is to specify which exponential family the response vector comes from. The corresponding summary results for this model are much the same as the linear mixed model, except for the estimated scale output which is used to detect overdispersion.

```
> summary(glmm)
Generalized linear mixed model fit using Laplace
Formula: BellbirdTotal ~ I(Year - 2003) * Treatment +(1 | fTransect/fRep/fCount)
```

```
 Family: poisson(log link)
  AIC  BIC logLik deviance
 1668 1706 -826.9     1654
Random effects:
 Groups                   Name        Variance   Std.Dev.
 fCount:(fRep:fTransect) (Intercept) 5.0000e-10 2.2361e-05
 fRep:fTransect          (Intercept) 1.5002e-03 3.8733e-02
 fTransect               (Intercept) 8.2180e-02 2.8667e-01
number of obs: 1800,
groups: fCount:(fRep:fTransect), 180; fRep:fTransect, 36; fTransect, 12

Estimated scale (compare to  1 )  0.8607026

Fixed effects:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                  0.43416    0.10312   4.210 2.55e-05 ***
I(Year - 2003)               0.01427    0.01825   0.782 0.434050
TreatmentTRUE               -0.31800    0.08803  -3.612 0.000303 ***
I(Year - 2003):TreatmentTRUE 0.12269    0.02552   4.807 1.53e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) I(Yr-2003) TrTRUE
I(Yer-2003) -0.536
TretmntTRUE -0.410  0.628
I(Y-2003):T  0.383 -0.715     -0.912
```

Using `lmer()` with respect to Poisson data allows there to now be p-values present. This therefore makes the nlme package somewhat unnecessary to use over the lme4 package.
Another factor which makes `lmer()` superior is the ability to modify the method in how the GLMM is modeled. GLMM estimates cannot be carried out in the same way as the linear mixed models as they require the evaluation of high dimensional integrals that cannot be done analytically, and so, approximation methods must be used. R possesses three of these methods. The default method is the Laplacian Approximation `"Laplace"`, and the other alternatives are the penalised quasi-likelihood `"PQL"` and the adaptive Gaussian

quadrature approximations "AGQ". The Laplacian Approximation method directly approximates the integrals using second order Taylor series expansions (Raudenbush and Bryk, 2002). Penalised quasi-likelihood estimates a quasi-likelihood function which only involves first and second order conditional moments (Molenberghs and Verbeke, 2005). The AGQ method is yet to be fully developed in R. The R help page on `lmer()` states that the Laplace method is more accurate than the PQL method so it is advisable to utilise the Laplace default method. Also, PQL methods bring about estimators which are asymptotically biased (McCulloch and Searle, 2001). It is , however, difficult to determine the quantity and therefore the impact that this bias brings. The PQL method can be implemented like so,

```
> glmm1<-lmer(BellbirdTotal~I(Year-2003)*Treatment + (1|fTransect/fRep/fCount),
        family=poisson, method="PQL")
```

The results of the above GLMM model on the bellbirds come out to be rather similar to the linear mixed model with *Year* still not showing any significance. By looking at the variance components in the random effects we can see that *Transect* is producing the majority of the variation in comparison to the *Rep* and *Count* random effects. The estimated scale gives a result of 0.86 which is reasonably close to one. This clarifies that the Poisson distribution was indeed appropriate as there is no overdispersion. Also, note the large decrease in the AIC when using the GLMM model to the linear mixed model, this confirms the model is far more appropriate.
If for some reason, it is necessary to apply GLMM's using the nlme package, it can be done in this way. Firstly another package (MASS) must be loaded which works by making repeated internal calls to the `lme()` function in the nlme package (Bivand et al., 2008).

```
> library(MASS)
> glmm2<-glmmPQL(BellbirdTotal~I(Year-2003)*Treatment,  random=~1|fTransect/fRep/fCount,
        family=poisson)
```

Obviously the `glmmPQL()` function only performs the PQL method, which is why it is at disadvantage to the `lmer()` function. However, just like with the `lme()` there are more functions that can be applied than in the lme4 package. These functions (such as `predict()`, `residuals()`, etc.) are inappropriate as they are using the internal calls to the `lme()` function which assumes it is a linear mixed model. Additionally, the `glmmPQL()` command does not give the log likelihood of the model and thus no AIC or BIC values are given.

```
> summary(glmm2)
Linear mixed-effects model fit by maximum likelihood
 Data: NULL
  AIC BIC logLik
```

```
   NA  NA      NA

Random effects:
 Formula: ~1 | fTransect
        (Intercept)
StdDev:    0.286762

 Formula: ~1 | fRep %in% fTransect
        (Intercept)
StdDev:  0.04535653

 Formula: ~1 | fCount %in% fRep %in% fTransect
        (Intercept)  Residual
StdDev:   0.1217458 0.8453115

Variance function:
 Structure: fixed weights
 Formula: ~invwt
Fixed effects: BellbirdTotal ~ I(Year - 2003) * Treatment
                              Value  Std.Error   DF   t-value p-value
(Intercept)                 0.4299996 0.09844648 1617  4.367852   0.000
I(Year - 2003)              0.0146485 0.01544212 1617  0.948607   0.343
TreatmentTRUE              -0.3163846 0.07450095 1617 -4.246718   0.000
I(Year - 2003):TreatTRUE    0.1222496 0.02159748 1617  5.660365   0.000
 Correlation:
                             (Intr) I(Yr-2003) TrTRUE
I(Year - 2003)               -0.475
TreatmentTRUE                -0.363  0.628
I(Year - 2003):TreatmentTRUE  0.340 -0.715     -0.913

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-1.96227211 -0.75331697 -0.03703465  0.62363962  3.95138795
```

```
Number of Observations: 1800
Number of Groups:
fTransect          fRep %in% fTransect fCount %in% fRep %in% fTransect
    12                                36                             180
```

The `glmmPQL()` output is reasonably similar to the `lmer()` except for a different layout of the random effects and the additional five number summary of the standardized within-group residuals.

Additional Analysis on Generalised Linear Mixed Models

Model Comparisons

Commonly, after the model is fitted, it is necessary to compare with other models in order to choose the most appropriate model for the data. This can be done with `lmer()` fitted models using the `anova()` function. For example, let us compare the two models above, `lmm` and `glmm`.

```
> anova(lmm,glmm)
Data:
Models:
lmm: BellbirdTotal ~ I(Year - 2003) * Treatment + (1 | fTransect)
glmm: BellbirdTotal ~ I(Year - 2003) * Treatment + (1 | fTransect/fRep/fCount)
     Df     AIC     BIC  logLik Chisq Chi Df Pr(>Chisq)
lmm   5  5558.9  5586.3 -2774.4
glmm  7  1667.8  1706.3  -826.9  3895      2  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 '
```

As you can see, the p-value is highly significant, indicating the more complex model `glmm` is better. Similarly, comparing model `glmm` with the same model (`glmm3`) except with *Reps* and *Counts* removed gives;

```
> glmm3<-lmer(BellbirdTotal~I(Year-2003)*Treatment + (1|fTransect), family=poisson)
> anova(glmm,glmm3)
Data:
Models:
glmm3: BellbirdTotal ~ I(Year - 2003) * Treatment + (1 | fTransect)
glmm: BellbirdTotal ~ I(Year - 2003) * Treatment + (1|fTransect/fRep/fCount)
      Df     AIC     BIC  logLik Chisq Chi Df Pr(>Chisq)
```

```
glmm3  5 1664.00 1691.47 -827.00
glmm   7 1667.81 1706.28 -826.91 0.183      2      0.9126
```

The p-value is not significant, showing that the less complex `glmm3` is a better model. This means that the *Rep* and *Counts* are not particularly informative in the model.
Model comparisons using `anova()` work fine when comparing two `lmer()` models, but unfortunately do not work when comparing with `lm()` or `glm()` models. Also, the `glmmPQL()` function does not work with model comparisons using `anova()` due to the PQL method being used. The PQL method however, when specified in the `lmer()` function does allow `anova()` comparisons.

Applying Random Slopes to a Mixed Model

In fitting our mixed models above, we were only modeling random intercepts. Modeling solely random slopes (not random intercepts) or even both random intercepts and random slopes is not that much different. Consider our model `glmm3` above, we fit the single random effect *Transect* with random intercepts. This can be remodeled with random slopes across the *Year* explanatory variable like so:

```
> glmm4<-lmer(BellbirdTotal~I(Year-2003)*Treatment + (I(Year-2003)-1|fTransect),
        family=poisson)
```

And with both random intercepts and random slopes like so:

```
> glmm5<-lmer(BellbirdTotal~I(Year-2003)*Treatment + (I(Year-2003)|fTransect),
        family=poisson)
```

This model `glmm5` gives two random terms, shown in the summary output below,

```
> summary(glmm5)
Generalized linear mixed model fit using Laplace
Formula: BellbirdTotal ~ I(Year - 2003) * Treatment + (I(Year - 2003) | fTransect)
 Family: poisson(log link)
  AIC  BIC logLik deviance
 1666 1705 -826.2     1652
Random effects:
 Groups    Name            Variance Std.Dev. Corr
 fTransect (Intercept)     0.057829 0.240478
           I(Year - 2003) 0.001085 0.032939 0.298
number of obs: 1800, groups: fTransect, 12
```

```
Estimated scale (compare to  1 )  0.8603476

Fixed effects:
                             Estimate Std. Error z value Pr(>|z|)
(Intercept)                   0.44411    0.09259   4.796 1.62e-06 ***
I(Year - 2003)                0.01025    0.02064   0.497 0.619373
TreatmentTRUE                -0.31623    0.08803  -3.592 0.000328 ***
I(Year - 2003):TreatmentTRUE  0.12220    0.02552   4.789 1.68e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) I(Yr-2003) TrTRUE
I(Yer-2003) -0.428
TretmntTRUE -0.456  0.555
I(Y-2003):T  0.427 -0.632     -0.912
```

The random effect variance for the intercept is shown with the variance of the random slopes shown directly under it. More variation is supplied by the intercepts as opposed to the slope.
To observe the exact coefficients for each level of the random effects the `coef()` function can be used. In our example, it is showing the unique intercept and slope for each of the twelve transects.

```
> coef(glmm5)
An object of class "coef.lmer"
[[1]]
   (Intercept) I(Year - 2003) TreatmentTRUE I(Year - 2003):TreatmentTRUE
1   -0.0865066   -0.054175399     -0.316229                    0.1222028
2    0.1818510    0.002892005     -0.316229                    0.1222028
3    0.6399849    0.005410279     -0.316229                    0.1222028
4    0.6753530    0.024612258     -0.316229                    0.1222028
5    0.3266931    0.039284875     -0.316229                    0.1222028
6    0.5674877    0.023167953     -0.316229                    0.1222028
7    0.6972835    0.018984672     -0.316229                    0.1222028
8    0.5172862    0.010263141     -0.316229                    0.1222028
9    0.6297028    0.023099662     -0.316229                    0.1222028
```

10	0.3670774	0.013613861	-0.316229	0.1222028
11	0.3284215	-0.013945011	-0.316229	0.1222028
12	0.5071552	0.031819649	-0.316229	0.1222028

The `coef()` function can be used across all mixed models in both packages.
With the undergoing developments of mixed modeling in R, future tools and functions will continue to be added to both packages, and thus continually improve the ability to analyse mixed models.

Conclusions for the Example Data

The purpose of the study was to illustrate the use of GLMM's in R. Analysis was done on nine species of birds, with only the bellbird model shown in this report. Full results and graphical summaries for these other species are not shown in this report but are available. Although all species showed varying results, there was evidence suggesting a positive effect on the number of birds due to the habitat restoration project. Individual bird species showed positive, negative or no trend at all (like the bellbirds) over the five years. If there was evidence in differences between the treatments it suggested that the treatment area has less birds. However, the interaction term was often showing higher increases in bird numbers in the treatment area which proved the project is showing promising signs. In order to establish more substantial evidence as to the effectiveness of the project, further counts will need to be obtained in future years. Five years of monitoring has not produced very comprehensive conclusions and so the overall trends of the bird numbers is still uncertain. The results of the analysis showed that the need for repeated counts within the day and among the weeks is questionable as opposed to the multiple transects which did supply useful information to the model. A more thorough analysis of the survey design would be needed before any survey-changes were introduced.

References

Bivand, R. S., Pebesma, E. J., and Gomez-Rubio, V. (2008). *Applied Spatial Data Analysis with R.* Springer, New York.

Faraway, J. J. (2006). *Extending the Linear Model with R.* Chapman and Hall/CRC, Boca Raton, Florida.

Fox, J. (2002). *An R and S-PLUS Companion to Applied Regression.* Sage Publications, Inc.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press, New York.

Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications.* Springer, New York.

Lee, Y., Nelder, J. A., and Pawitan, Y. (2006). *Generalized Linear Models with Random Effects.* Chapman and Hall/CRC, Boca Raton, Florida.

McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear, and Mixed Models.* John Wiley and Sons, Inc, Canada.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data.* Springer, New York.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS.* Springer, New York.

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models.* Sage Publications, Inc., California.